

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2 0 0 3 年 1 0 月 1 4 日

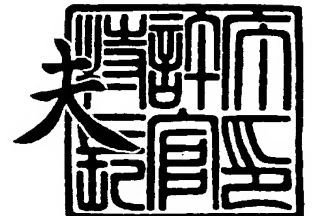
出 願 番 号
Application Number: 特 願 2 0 0 3 - 3 5 3 2 1 2
[ST. 10/C]: [J P 2 0 0 3 - 3 5 3 2 1 2]

出 願 人
Applicant(s): 株式会社日立製作所

2 0 0 4 年 2 月 1 9 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫





【書類名】 特許願
【整理番号】 NT03P0795
【提出日】 平成15年10月14日
【あて先】 特許庁長官 殿
【国際特許分類】 G06F 12/00
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作
 所 システム開発研究所内
 【氏名】 藤原 啓成
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作
 所 システム開発研究所内
 【氏名】 岩見 直子
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作
 所 システム開発研究所内
 【氏名】 渡辺 直企
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作
 所 システム開発研究所内
 【氏名】 志賀 賢太
【特許出願人】
 【識別番号】 000005108
 【氏名又は名称】 株式会社日立製作所
【代理人】
 【識別番号】 100068504
 【弁理士】
 【氏名又は名称】 小川 勝男
 【電話番号】 03-3661-0071
【選任した代理人】
 【識別番号】 100086656
 【弁理士】
 【氏名又は名称】 田中 恭助
 【電話番号】 03-3661-0071
【選任した代理人】
 【識別番号】 100094352
 【弁理士】
 【氏名又は名称】 佐々木 孝
 【電話番号】 03-3661-0071
【手数料の表示】
 【予納台帳番号】 081423
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1

【書類名】 特許請求の範囲**【請求項 1】**

ネットワークを介してホスト計算機および他のストレージ装置と通信するストレージ装置であって、前記ストレージ装置は、複数のメモリバッファから成る使用可能バッファと、通信のために割り当てられた複数のメモリバッファから成る使用中バッファと、外部からの照会に応答して前記使用可能バッファのサイズを通知する手段と、バッファ予約対象と予約バッファサイズを指定する外部からのバッファ予約要求に応答して前記使用可能バッファから前記予約バッファサイズ分のメモリバッファを取り出して前記バッファ予約対象のために予約バッファとして確保する手段と、外部からのアプリケーション開始要求に応答して前記予約バッファ中のメモリバッファを割り当てて前記使用中バッファとする手段と、割り当てられた前記使用中バッファを用いて通信するアプリケーションを実行する手段とを有することを特徴とするストレージ装置。

【請求項 2】

前記予約バッファサイズは、1つのコネクションについて単位時間当たりのデータ転送量である帯域にデータ送信してからその送達確認を受信するまでの遅延時間に乗じた数値であることを特徴とする請求項 1 に記載のストレージ装置。

【請求項 3】

前記ストレージ装置は、前記バッファ予約対象と変更後の予約バッファサイズを指定する外部からの予約バッファ更新要求に応答して元の前記予約バッファサイズを拡大する場合には前記使用可能バッファから増加サイズ分のメモリバッファを取り出して元の前記予約バッファサイズに追加する手段と、元の前記予約バッファサイズを縮小する場合には元の前記予約バッファから減少サイズ分のメモリバッファを取り出して前記使用可能バッファに戻す手段とをさらに有することを特徴とする請求項 1 に記載のストレージ装置。

【請求項 4】

前記アプリケーションは、当該ストレージ装置に格納されるデータを他のストレージ装置にコピーするリモートコピー処理であることを特徴とする請求項 1 に記載のストレージ装置。

【請求項 5】

前記バッファ予約対象は、当該ストレージ装置の通信ポートと、前記他のストレージ装置の通信ポートとの間のポート間通信であることを特徴とする請求項 1 または 4 に記載のストレージ装置。

【請求項 6】

前記バッファ予約対象は、当該ストレージ装置の通信ポートと前記他のストレージ装置の通信ポートとの組であるポートペアおよび前記他のストレージ装置側の前記アプリケーションが適用するプロトコルを示す番号の組合せであることを特徴とする請求項 1 または 4 に記載のストレージ装置。

【請求項 7】

前記バッファ予約対象は、当該ストレージ装置と通信相手装置との間のコネクションであることを特徴とする請求項 1 に記載のストレージ装置。

【請求項 8】

前記ストレージ装置は、ディスクから読み出したデータを一時記憶するキャッシュメモリを備え、外部からの照会に応答して前記キャッシュメモリの空き容量を通知する手段と、前記バッファ予約対象と変更後の予約バッファサイズを指定する外部からの予約バッファ拡大要求に応答して前記使用可能バッファサイズから増加サイズ分のメモリバッファを取り出して元の前記予約バッファサイズに追加する手段とをさらに有することを特徴とする請求項 1 に記載のストレージ装置。

【請求項 9】

前記バッファは、TCP を用いて送信したデータを記憶するバッファであることを特徴とする請求項 1 に記載のストレージ装置。

【請求項 10】

前記他のストレージ装置との間の通信について通信プロトコルを処理する手段は、前記予約バッファサイズを輻輳ウィンドウの最大値とすることを特徴とする請求項 9 に記載のストレージ装置。

【請求項 1 1】

前記バッファは、ファイバ・チャネル・プロトコルを用いて送信したデータを記憶するクレジットであることを特徴とする請求項 1 に記載のストレージ装置。

【請求項 1 2】

前記予約バッファとして確保する手段は、使用可能なバッファが不足する場合、メモリからバッファとして使用可能な領域を割り当ててことを特徴とする請求項 1 に記載のストレージ装置。

【請求項 1 3】

ネットワークを介してホスト計算機および他のストレージ装置と通信するストレージ装置であって、前記ストレージ装置は、複数のメモリバッファから成る使用可能バッファと、通信のために割り当てられた複数のメモリバッファから成る使用中バッファと、外部からの照会に 응답して前記使用可能バッファのサイズを通知する通信インタフェース情報取得部と、バッファ予約対象と予約バッファサイズを指定する外部からのバッファ予約要求に 응답して前記使用可能バッファから前記予約バッファサイズ分のメモリバッファを取り出して前記バッファ予約対象のために予約バッファとして確保し、外部からのアプリケーション開始要求に 응답して前記予約バッファ中のメモリバッファを割り当てて前記使用中バッファとするバッファ制御部と、割り当てられた前記使用中バッファを用いて通信するアプリケーションを実行するアプリケーション実行部とを有することを特徴とするストレージ装置。

【請求項 1 4】

ネットワークを介してホスト計算機および他のストレージ装置と通信するストレージ装置と、前記ネットワークを介して前記ストレージ装置と通信するストレージ管理装置とを有するシステムであって、

前記ストレージ装置は、複数のメモリバッファから成る使用可能バッファと、通信のために割り当てられた複数のメモリバッファから成る使用中バッファと、前記ストレージ管理装置からの照会に 응답して前記使用可能バッファのサイズを通知する手段と、バッファ予約対象と予約バッファサイズを指定する前記ストレージ管理装置からのバッファ予約要求に 응답して前記使用可能バッファから前記予約バッファサイズ分のメモリバッファを取り出して前記バッファ予約対象のために予約バッファとして確保する手段と、前記ストレージ管理装置からのアプリケーション開始要求に 응답して前記予約バッファ中のメモリバッファを割り当てて前記使用中バッファとする手段と、割り当てられた前記使用中バッファを用いて通信するアプリケーションを実行する手段とを有し、

前記ストレージ管理装置は、前記ストレージ装置に前記使用可能バッファのサイズを照会する手段と、前記ストレージ装置に前記バッファ予約要求を送信する手段と、前記ストレージ装置に前記アプリケーション開始要求を送信する手段とを有することを特徴とするストレージシステム。

【請求項 1 5】

前記予約バッファサイズは、1つのコネクションについて単位時間当たりのデータ転送量である帯域にデータ送信してからその送達確認を受信するまでの遅延時間に乗じた数値であることを特徴とする請求項 1 4 に記載のストレージシステム。

【請求項 1 6】

前記ストレージ装置は、前記バッファ予約対象と変更後の予約バッファサイズを指定する前記ストレージ管理装置からの予約バッファ更新要求に 응답して元の前記予約バッファサイズを拡大する場合には前記使用可能バッファから増加サイズ分のメモリバッファを取り出して元の前記予約バッファサイズに追加する手段と、元の前記予約バッファサイズを縮小する場合には元の前記予約バッファから減少サイズ分のメモリバッファを取り出して前記使用可能バッファに戻す手段とをさらに有し、

前記ストレージ管理装置は、前記予約バッファ更新要求を送信する手段をさらに有することを特徴とする請求項 14 に記載のストレージシステム。

【請求項 17】

前記アプリケーションは、当該ストレージ装置に格納されるデータを他のストレージ装置にコピーするリモートコピー処理であることを特徴とする請求項 14 に記載のストレージシステム。

【請求項 18】

前記バッファ予約対象は、当該ストレージ装置の通信ポートと、前記他のストレージ装置の通信ポートとの間のポート間通信であることを特徴とする請求項 14 または 17 に記載のストレージシステム。

【請求項 19】

前記バッファ予約対象は、当該ストレージ装置の通信ポートと前記他のストレージ装置の通信ポートとの組であるポートペアおよび前記他のストレージ装置側の前記アプリケーションが適用するプロトコルを示す番号の組合せであることを特徴とする請求項 14 または 17 に記載のストレージシステム。

【請求項 20】

前記バッファ予約対象は、当該ストレージ装置と通信相手装置との間のコネクションであることを特徴とする請求項 14 に記載のストレージシステム。

【請求項 21】

前記ストレージ装置は、ディスクから読み出したデータを一時記憶するキャッシュメモリを備え、前記ストレージ管理装置からの照会に回答して前記キャッシュメモリの空き容量を通知する手段と、前記バッファ予約対象と変更後の予約バッファサイズを指定する前記ストレージ管理装置からの予約バッファ拡大要求に回答して前記使用可能バッファサイズから増加サイズ分のメモリバッファを取り出して元の前記予約バッファサイズに追加する手段とをさらに有し、

前記ストレージ管理装置は、前記ストレージ装置に前記キャッシュメモリの空き容量を照会する手段と、前記空き容量が所定値を下回ったとき前記予約バッファ拡大要求を送信する手段とをさらに有することを特徴とする請求項 14 に記載のストレージシステム。

【請求項 22】

前記ストレージ管理装置は、さらに前記バッファ予約対象の種別を選択可能な表示画面を表示する手段を備えることを特徴とする請求項 14 に記載のストレージシステム。

【請求項 23】

前記ストレージ装置は、さらに前記予約バッファサイズを計算するバッファサイズ算出手段を備えることを特徴とする請求項 14 に記載のストレージシステム。

【請求項 24】

前記ストレージシステムには、通信路の利用帯域や遅延を計測する手段を有する計測装置が接続されており、

前記ストレージ管理装置の前記バッファサイズ算出手段は、前記計測装置から通信路の利用帯域や遅延情報を受信し、それらのデータに基づいて前記予約バッファサイズを計算することを特徴とする請求項 23 に記載のストレージシステム。

【請求項 25】

前記ストレージ管理装置は、前記計測装置から受信した通信路の利用帯域や遅延を記憶する通信路データ記憶手段を備えることを特徴とする請求項 24 に記載のストレージシステム。

【書類名】明細書**【発明の名称】**通信バッファ予約機能を備えるストレージ装置およびシステム**【技術分野】****【0 0 0 1】**

本発明は、ネットワークに接続され、複数の情報処理装置とネットワークを介して通信するストレージ装置に係わり、特に通信インタフェースのバッファを制御する機能を備えるストレージ装置に関する。

【背景技術】**【0 0 0 2】**

企業および組織の基幹業務へ情報技術が浸透するのに伴い、企業および組織が保有する電子データの量とその重要性が増している。このような企業および組織が、地震、停電、テロなどの大規模災害が発生した場合にも、保有するデータを保護でき、かつ迅速に修復できるためには、データの正確なコピーを保持することが必要である。

【0 0 0 3】

通常、このようなデータのコピーは、災害の被害を局所化するため、地理的に離れた（たとえば数百キロメートル離れた）拠点間で行う。各拠点には、ホストとデータの記憶システム（たとえばストレージ装置）が設置され、拠点間は広域イーサネット（イーサネットは、富士ゼロックス株式会社の商品名称である）、あるいはインターネットなどの広域ネットワーク（WAN）で接続される。データの原本を記憶するストレージ装置をマスタストレージ装置、マスタストレージ装置と接続されたホストをマスタホスト、マスタストレージ装置やマスタホストが設置された拠点をマスタサイトという。一方、データのコピー先のストレージ装置をリモートストレージ装置、リモートストレージ装置と接続されたホストをリモートホスト、リモートストレージ装置やリモートホストが設置された拠点をリモートサイトという。このように地理的に離れた拠点間でデータをコピーすることをリモートコピーといい、リモートコピーを実行するシステムをリモートコピーシステムという。さらにリモートコピーを行うマスタストレージ装置の通信ポートをマスタポート、リモートストレージ装置の通信ポートをリモートポートといい、マスタポートのアドレスとリモートポートのアドレスの組をポートペアという。

【0 0 0 4】

リモートコピーには、同期リモートコピーと非同期リモートコピーがある。同期リモートコピーでは、マスタホストの書き込み処理は、書き込みデータをリモートコピーした後完了する。一方、非同期リモートコピーでは、マスタホストの書き込み処理は、マスタストレージ装置が書き込みデータをメモリに待避したとき直ちに完了する。その後、マスタストレージ装置がメモリに待避したデータをリモートコピーする。

【0 0 0 5】

通常、リモートコピーシステムの拠点間通信には、重要なデータに欠損が生じるのを防止するため、TCP/IPなどの信頼性を保証するプロトコルが用いられる。具体的には、送信側ノードが送信データのコピーを通信用メモリ領域に保持しておき、受信側ノードから送達確認通知を受信するまで、再送を繰り返すことによって信頼性を確保する。ここで通信用メモリ領域とは、システムがネットワークプロトコル専用に割り当てたページの集合である。ページとはメモリを固定長の単位で分割した記憶領域の単位である。システムは、メモリ管理機能によりページを固定長のバッファ用データ構造であるメモリバッファに分割し、使用可能なメモリバッファのリストである自由バッファリストに加える。システムは、バッファ割当要求を受けると、自由バッファリストから必要サイズ分のメモリバッファを取り出し、TCPコネクションなどの通信に割り当てる。自由バッファリストのサイズが不足した場合、システムは未分割なページを分割して自由バッファリストに加える。通信用メモリ領域に未分割なページが無い場合、システムは、システムのメモリ資源から新しくページを増やす。システムが利用可能なメモリ資源をもっていない場合、メモリ不足を表すエラーを返す。以降、自由バッファリストのサイズを、使用可能バッファサイズといい、また通信用メモリ領域のうち、送信データのコピーを保持するために通信

に割り当てるメモリバッファのリストを単にバッファという。

【0006】

このバッファに必要とされるサイズは、通信路の遅延および帯域に比例する。ここで遅延とは、送信側ノードがデータを送信してから送達確認通知を受信するまでの時間とする。遅延や帯域は経路によって異なるため、障害発生などの理由により通信中に経路が変化した場合、通信路の遅延や帯域も変化する。また遅延や帯域は、経路上のネットワーク機器の負荷や輻輳の有無によっても変化する。このように通信路の遅延や帯域は通信中に変化するため、結果的に必要なバッファサイズも通信中に変化する。

【0007】

もし通信路の変化に関わらずバッファサイズが一定の場合、通信性能の劣化という問題が生じる。すなわち送信側ノードが送達確認通知を待たずに送信できるデータ量は、バッファの未使用領域のサイズを限度とするため、バッファサイズが過小な場合、通信効率が低下する。逆にバッファサイズが過大であり、送信側ノードが通信路のデータ転送能力を超えたデータを通信路に送信すると、経路上のネットワーク機器がデータを破棄することがある。データの破棄が発生すると、送信側ノードがデータを再送するため、結果的に通信効率が低下する。したがって、通信路の遅延や帯域の変化に応じてバッファサイズを変更する技術が重要になる。

【0008】

通信路の遅延や帯域の変化に応じてバッファサイズを変更する技術は、たとえば特許文献1に記載されている。この技術は、コネクションごとに割り当てた事前割振りバッファのサイズを、事前割振りバッファの使用率に応じて変更する。すなわち、コネクションの生成時に、そのコネクションで使用する事前割振りバッファのサイズの最大値と最小値、および使用率の上限値と下限値を設定する。そして通信開始後、送受信動作ごと、および（または）定期的に事前割振りバッファの使用率を調べ、平均使用率を計算する。この平均使用率が上限値を超え、かつ事前割振りバッファのサイズが最大値を超えていない場合、バッファを拡大する。逆にこの平均使用率が下限値を下回り、かつ事前割振りバッファのサイズが最小値を超えていない場合、バッファを縮小する。通信路の遅延が増大すると事前割振りバッファの平均使用率が上がり、逆に遅延が減少すると平均使用率が下がるため、上記技術により、通信性能の低下を防ぐことができる。

【0009】

【特許文献1】特開2002-208981号公報

【発明の開示】

【発明が解決しようとする課題】

【0010】

上記従来技術では、通信用メモリ領域の不足により、必要なサイズのバッファを確保できない場合がある。これは大きなバッファを必要とする遅延が大きい通信路において特に問題である。問題が起こる例として、上記のリモートコピーシステムにおいて、マスタストレージ装置がリモートストレージ装置との間のリモートコピー通信と、マスタホストとの間のディスクI/O通信とを同じ通信インターフェースを用いて実行する場合が挙げられる。このような場合に、多数のホストがディスクI/O通信を実行すると、通信用メモリ領域が使い尽くされ、リモートコピーの開始時にリモートコピー通信用のバッファが確保できず、その結果、リモートコピーが中断してしまう。万が一、このような時に大規模災害が発生すると、重要なデータが失われ、企業および組織が多大な損害をこうむる。

【0011】

また遅延の変化が大きい通信路においても問題が発生する。例えば上記のリモートコピーシステムにおいて、マスタストレージ装置がリモートコピーを実施中に、マスタホストとの間のI/O数が増大し、通信用メモリ領域が使い尽くされたとする。その直後、リモートコピー用の通信路の遅延が増大しても、バッファを拡大できないため、リモートコピーのデータ送信が停止する。同期リモートコピーの場合、リモートコピーが停止するとマスタホストのディスクI/O処理も停止する。また、非同期リモートコピーの場合でも、

リモートコピーの停止が続くと、ストレージ装置のメモリが未送信データで使い尽くされるため、マスタホストのディスク I / O 処理が停止する。この結果、マスタホスト上のアプリケーションがディスク I / O エラーが発生したとみなし、基幹業務の処理が停止してしまう。

【0 0 1 2】

したがって、改善されたバッファ制御技術が求められている。

【課題を解決するための手段】

【0 0 1 3】

本発明は、ネットワークを介してホスト計算機および他のストレージ装置と通信するストレージ装置であって、複数のメモリバッファから成る使用可能バッファと、通信のために割り当てられた複数のメモリバッファから成る使用中バッファと、外部からの照会に回答して使用可能バッファのサイズを通知する手段と、バッファ予約対象と予約バッファサイズを指定する外部からのバッファ予約要求に回答して使用可能バッファから指定された予約バッファサイズ分のメモリバッファを取り出して指定されたバッファ予約対象のために予約バッファとして確保する手段と、外部からのアプリケーション開始要求に回答して予約バッファ中のメモリバッファを割り当てて使用中バッファとする手段と、割り当てられた使用中バッファを用いて通信するアプリケーションを実行する手段とを有するストレージ装置を特徴とする。

【発明の効果】

【0 0 1 4】

本発明によれば、あらかじめ予約バッファを確保するので、バッファ予約対象を用いて行う通信のバッファ不足を回避することができ、アプリケーション処理の中断を防止できる。

【発明を実施するための最良の形態】

【0 0 1 5】

以下、本発明の実施例について図面を用いて説明する。以下の図中、同一の部分には同一の符号を付加する。

【0 0 1 6】

図 1 は、第 1 の実施例のリモートコピーシステムの構成図である。本システムは、マスタサイトとリモートサイトを有する。マスタサイトは、マスタストレージ装置 1、通信路の帯域、および遅延などの通信路情報を計測する通信路情報取得装置 4、通信路情報取得装置 4、5 から計測した通信路情報を収集する通信路情報管理装置 3、マスタホスト 6 およびマスタストレージ装置 1 の管理端末であるストレージ管理装置 8 の構成機器を有する。これらの構成機器は、ギガビットイーサネットなどの LAN 2 1 9 を介して接続される。リモートサイトは、リモートストレージ装置 2、通信路情報取得装置 5 およびリモートホスト 7 が LAN 2 2 0 を介して接続される。マスタサイトとリモートサイトとは、WAN 2 2 1 を介して接続される。

【0 0 1 7】

マスタストレージ装置 1 は、ハードディスクなどの二次記憶装置（以下、ディスクという）2 0 0、2 0 1、2 0 2、これらディスクへの入出力を制御するディスクコントローラ 2 0 5、およびディスクとディスクコントローラ 2 0 5 とを接続するバスなどの通信線である入出力（I/O）バス 2 0 4 から構成される。

【0 0 1 8】

さらにディスクコントローラ 2 0 5 は、LAN 2 1 9 を経由して他装置と通信する通信インタフェース（I/F）2 0 8、主記憶装置（以下、メモリという）2 1 8、中央演算装置（以下、CPU という）2 1 7、I/O バス 2 0 4 と接続する I/O インタフェース 2 0 6、2 0 7、およびバスなどの通信線（以下、内部バスという）2 1 6 から構成される。メモリ 2 1 8 は、ディスクから読み出したデータを一時記憶するキャッシュメモリ 1 0 2、およびリモートコピーを実行するアプリケーションプログラムであるリモートコピー部 1 0 3 を格納する。このリモートコピー部 1 0 3 は、あらかじめ、または可搬型記録媒体

からの読み込み、または他装置からのLAN 219経由のダウンロードにより、ディスクに格納され、必要に応じてメモリ 218に転送された後、CPU 217によって実行される。

【0019】

通信インタフェース 208は、バッファとして使用するメモリ領域である通信用メモリ領域 100、予約済みバッファを管理する予約バッファ情報テーブル 101、および割当済のバッファを管理する割当バッファ情報テーブル 109を格納するメモリ 209を有する。また通信インタフェース 208は、通信用メモリ領域 100のうち未使用領域のサイズを調べる通信インタフェース情報取得部 210、バッファの予約を行うバッファ制御部 215および他装置との通信を処理する通信プロトコル処理部 214のハードウェア又はプログラムを有する。また通信インタフェース 208は、LAN 219を構成するケーブルを接続するポート 211、212、213を有する。通信用メモリ領域 100の詳細については後述する。

【0020】

ストレージ管理装置 8は、通信インタフェース 222、メモリ 223、ディスク 224、CPU 226、表示装置などの出力装置（以下、ディスプレイという）227、キーボードなどの文字入力装置 228、マウス、タッチパネルなどのポインティング装置 229、および内部バス 225から構成される。メモリ 223は、通信路情報管理装置 3やマスタストレージ装置 1から得た通信路情報や通信インタフェース情報をディスク 224内のテーブルに記憶する情報管理部 104、ポートペア間の通信で使用するバッファのサイズを算出するバッファサイズ算出部 105、バッファサイズをマスタストレージ装置 1の通信インタフェース 209へ送信するバッファサイズ送信部 106、およびシステム管理者が作業を行うGUIを表示するGUI制御部 107を格納する。これらのプログラムは、あらかじめ、または可搬型記録媒体からの読み込み、または他装置からのLAN 219経由のダウンロードにより、ディスク 224に格納され、必要に応じてメモリ 223に転送された後、CPU 226によって実行される。ディスク 224は、ポートペア間の通信路の各種情報を記憶する通信路情報テーブル 108を格納する。このテーブルはメモリ 223に格納してもよい。

【0021】

リモートストレージ装置 2は、基本的にマスタストレージ装置 1と同一の構成機器およびプログラムを有する。マスタホスト 6、およびリモートホスト 7は、サーバ計算機やメインフレームなどの従来のホスト計算機である。

【0022】

図2は、通信用メモリ領域 100の構成を示す図である。通信用メモリ領域 100は、通常バッファ領域 31と予約バッファ領域 30から構成される。通常バッファ領域 31は、従来からあるバッファ領域であり、バッファ 34と自由バッファリスト 35から成る。自由バッファリスト 35は、使用可能な複数のメモリバッファのプールである。バッファ 34は、使用中の複数のメモリバッファの集合である。予約バッファ領域 30は、本発明のために追加されたバッファ領域であり、バッファ 34と予約バッファリスト 33から成る。予約バッファリスト 33は、予約された複数のメモリバッファであり、ポートペアごとに設けられる。バッファ 34は、通信に使用するための割当済の複数のメモリバッファの集合である。予約バッファリスト 33中のメモリバッファが使用のために割り当てられると、バッファ 34となる。自由バッファリスト 35中のメモリバッファが使用されると、通常バッファ領域 31内のバッファ 34又は予約バッファリスト 33に変わる。従って通常バッファ領域 31と予約バッファ領域 30との間に仕切りがあるわけではない。

【0023】

予約バッファリスト 33、自由バッファリスト 35およびバッファ 34は、すべて同一サイズの固定長メモリバッファから構成される。各メモリバッファは、アドレス部、データ部およびチェイン部から構成される。アドレス部は、データ部の先頭のメモリアドレスを格納する。データ部は、送受信データ部を格納する。チェイン部は、チェインによって

つながれた次のメモリバッファのメモリアドレスを格納する。メモリバッファは、ポートペアごとにバッファチェインを構成する。

【0024】

通信用メモリ領域100に設けられるバッファは、ファイバ・チャネル・プロトコルを用いて送信したデータを記憶するクレジットであってもよい。

【0025】

図3(a)は、予約バッファ情報テーブル101のデータ構成を示す図である。予約バッファ情報テーブル101は、予約バッファについての情報を記憶する。予約バッファ情報テーブル101の各レコードは、マスタポートIPアドレス301、リモートポートIPアドレス302、バッファアドレス303および予約バッファサイズ304から構成される。マスタポートIPアドレス301はマスタポートのIPアドレスであり、リモートポートIPアドレス302はリモートポートのIPアドレスである。これらマスタポートとリモートポートはポートペアを構成する。バッファアドレス303は、当該ポートペアについて予約した予約バッファリスト33の先頭メモリバッファのアドレスを格納する。予約バッファサイズ304は、予約されたバッファの容量、すなわち当該予約バッファリスト33に含まれるメモリバッファのサイズの合計値である。例えばレコード305は、マスタポートIPアドレス「211. 19. 1. 20」とリモートポートIPアドレス「133. 185. 12. 20」から成るポートペアに対して予約した予約バッファリスト33のアドレスが、「0x0FFFFFFF」であり、予約バッファのサイズが2. 1Mバイトであることを表す。

【0026】

なお図示しないが、割当バッファ情報テーブル109の各レコードは、予約バッファ情報テーブル101と同様に、マスタポートIPアドレス、リモートポートIPアドレス、バッファアドレスおよび割当バッファサイズから構成される。バッファアドレスは、バッファ34の先頭メモリバッファのアドレスである。割当バッファサイズは、バッファ34に含まれるメモリバッファのサイズの合計値である。

【0027】

図3(b)は、通信路情報テーブル108のデータ構成を示す図である。通信路情報テーブル108は、ポートペア間の通信路についての情報を記憶する。ここで通信路情報とは、ポートペア間の通信路の帯域、遅延、最大セッション数、および1セッション当たりのコネクション数のことをいう。通信路情報テーブル108の各レコードは、マスタポートIPアドレス401、リモートポートIPアドレス402、帯域403、遅延404、セッション数405およびコネクション数406から構成される。マスタポートIPアドレス401で示されるマスタポートとリモートポートIPアドレス402で示されるリモートポートはポートペアを構成する。帯域403は、当該ポートペア間で転送されるデータの単位時間あたりの転送量である。遅延404は、当該ポートペア間でデータ送信してからその送達確認を受信するまでの時間である。セッション数405は、当該ポートペア間についてリモートコピー部103が同時に実行するセッション数である。コネクション数406は、当該ポートペア間について1セッション当たりのTCPコネクションの数である。例えばレコード407は、マスタポートIPアドレス「211. 19. 1. 20」とリモートポートIPアドレス「133. 185. 12. 20」から成るポートペアにおいて、利用可能な帯域が「10M」bps、通信路をデータが往復するのに要する遅延時間が「2. 8μ」秒、セッション数が「2」、1セッションあたりのコネクション数が「3」であることを表す。

【0028】

次に図4を用いて、本実施例に係わるグラフィカルユーザインタフェース(GUI)について説明する。

【0029】

図4(a)は、リモートコピーシステムの管理者が、ポートペアに関する情報を登録するためのポートペア入力画面601の表示例を示す。ストレージ管理装置8のGUI制御

部 1 0 7 は、ディスプレイ 2 2 7 にポートペア入力画面 6 0 1 を表示する。システム管理者は、この画面を介し、文字入力装置 2 2 8 およびポインティング装置 2 2 9 を用いて各パラメータを設定する。

【 0 0 3 0 】

ポートペア入力画面 6 0 1 は、ポートペアのうちマスタポートの I P アドレスを入力する領域 6 0 2、ポートペアのうちリモートポートの I P アドレスを入力する領域 6 0 3、ポートペアあたりのリモートコピーセッションの最大数を入力する領域 6 0 4 および 1 セッションあたりの T C P コネクション数を入力する領域 6 0 5 を設ける。またポートペア入力画面 6 0 1 は、これらのパラメータから成るポートペア情報の登録を指示するための O K ボタン 6 0 6 およびポートペアの情報の入力を取り消すよう指示するためのキャンセルボタン 6 0 7 を有する。

【 0 0 3 1 】

図 4 (b) は、システム管理者が、特定のポートペアについて予約するバッファサイズを指定するためのサイズ指定画面 7 0 1 の表示例を示す。ストレージ管理装置 8 の G U I 制御部 1 0 7 は、ディスプレイ 2 2 7 にサイズ指定画面 7 0 1 を表示する。システム管理者は、この画面を介し、文字入力装置 2 2 8 およびポインティング装置 2 2 9 を用いて各パラメータを設定する。

【 0 0 3 2 】

サイズ指定画面 7 0 1 は、使用可能なバッファサイズを表示する領域 7 0 2、後述する処理により算出された推奨バッファサイズを表示する領域 7 0 3、バッファサイズの自動指定を選択するボタン 7 0 4、バッファサイズの手動指定を選択するボタン 7 0 5 およびバッファサイズを手動指定する場合にバッファサイズを入力する領域 7 0 6 を設ける。

【 0 0 3 3 】

ここで「使用可能なバッファサイズ」は、使用又は予約されていないバッファのサイズ、すなわち自由バッファリスト 3 5 に含まれるメモリバッファのサイズの合計値である。「推奨バッファサイズ」は、当該ポートペアについて推奨される予約バッファのサイズである。「予約バッファサイズ」は、当該ポートペアについて設定する予約バッファのサイズである。バッファサイズの「自動指定」が指示されると、ストレージ管理装置 8 は、推奨バッファサイズを予約バッファサイズとする。システム管理者が当該ポートペアの重要性に応じて予約バッファサイズを入力する場合には、「手動指定」が指示される。

【 0 0 3 4 】

またサイズ指定画面 7 0 1 は、指定したサイズのバッファを予約するよう指示するための O K ボタン 7 0 7 および予約するバッファサイズの入力を取り消すよう指示するためのキャンセルボタン 7 0 8 を有する。

【 0 0 3 5 】

図 5 は、システム管理者が、ストレージ管理装置 8 を介してリモートコピーのためのバッファを予約したのちに、マスタストレージ装置 1 からリモートストレージ装置 2 へのリモートコピーを開始するときの処理手順を示す。

【 0 0 3 6 】

最初にシステム管理者は、ストレージ管理装置 8 の G U I 制御部 1 0 7 を介してリモートコピー開始操作を行う。G U I 制御部 1 0 7 は、システム管理者からリモートコピー開始操作を受け付けると、ディスプレイ 2 2 7 にポートペア入力画面 6 0 1 を表示する。システム管理者が、文字入力装置 2 2 8、ポインティング装置 2 2 9 などを用いて、ポートペア情報を入力した後、O K ボタン 6 0 6 を押下すると、G U I 制御部 1 0 7 は、入力されたポートペア情報を情報管理部 1 0 4 へ通知する。情報管理部 1 0 4 は、このポートペア情報を受け取り、そのポートペア情報に基づき通信路情報テーブル 1 0 8 にレコードを追加する（ステップ 8 0 1）。

【 0 0 3 7 】

次に情報管理部 1 0 4 は、バッファの算出に必要な情報を収集する。まず情報管理部 1 0 4 は、通信路情報管理装置 3 へ特定のポートペアを指定する通信路情報取得要求を送信

する(ステップ802)。通信路情報管理装置3は、この通信路情報取得要求を受信すると、通信路情報取得装置4、5から、WAN221の帯域と遅延を取得する通信路情報計測処理を実行し(ステップ803)、その取得結果を含む通信路情報取得応答を情報管理部104へ送信する。情報管理部104は、この通信路情報取得応答を受信すると、それからポートペア間の通信路の帯域と遅延を取り出し、当該ポートペアのIPアドレスをキーに通信路情報テーブル108を検索し、ヒットしたレコードの帯域403と遅延404の値を設定する(ステップ804)。

【0038】

次に情報管理部104は、マスタストレージ装置1へ、当該ポートペアを指定する通信インタフェース情報取得要求を送信する(ステップ805)。マスタストレージ装置1の通信インタフェース情報取得部210は、この通信インタフェース情報取得要求を受信すると、通信インタフェース208の通信用メモリ領域100の使用可能バッファサイズを調べる通信インタフェース情報計測処理を実行し(ステップ806)、その結果得た使用可能バッファサイズを含む通信インタフェース情報取得応答をストレージ管理装置8へ送信する。ストレージ管理装置8の情報管理部104は、この通信インタフェース情報取得応答を受信すると、通信インタフェース情報取得応答から取り出した使用可能バッファサイズを、メモリ223上に保持する(ステップ807)。上記の通信路情報の取得手順(ステップ802、803、804)と通信インタフェース情報の取得手順(ステップ805、806、807)の順序は問わない。

【0039】

その後、バッファサイズ算出部105は、後述する推奨バッファサイズ算出処理を実行する(ステップ808)。GUI制御部107は、その実行結果を受けてディスプレイ227に先に説明したサイズ指定画面701を表示する。このサイズ指定画面701の表示領域702は、ステップ807で得た使用可能バッファサイズを表示する。また表示領域703は、ステップ808でバッファサイズ算出部105が算出した推奨バッファサイズを表示する。

【0040】

システム管理者がOKボタン707を押下すると、GUI制御部107は、バッファサイズ送信部106に予約バッファサイズを通知する(ステップ809)。ここでシステム管理者が自動指定ボタン704を選択した場合、取得した推奨バッファサイズが予約バッファサイズとされる。逆に手動指定ボタン705を選択した場合、システム管理者がバッファサイズ入力領域706に入力したバッファサイズを予約バッファサイズとする。ただしシステム管理者が、0以下、もしくは使用可能バッファサイズを越える値を入力した場合、GUI制御部107は、バッファサイズ入力領域706をクリアし、再入力を促す警告を発する。

【0041】

そしてバッファサイズ送信部106は、予約バッファサイズを受け取ると、この予約バッファサイズ、およびシステム管理者がポートペア入力画面601で入力したポートペアを含むバッファ予約要求を、マスタストレージ装置1へ送信する(ステップ810)。

【0042】

マスタストレージ装置1がこのバッファ予約要求を受信すると、バッファ制御部215は、後述するバッファ予約処理を行う(ステップ811)。バッファ制御部215は、バッファ予約処理を実行後、ストレージ管理装置8へバッファ予約応答を送信する(ステップ812)。

【0043】

ストレージ管理装置8がバッファ予約応答を受信すると、バッファ予約が完了する。情報管理部104は、バッファ予約が完了した後、各セッションのコネクションごとに当該ポートペアを指定するリモートコピー開始要求を、リモートコピー部103へ通知する(ステップ813)。なおこの際に情報管理部104は、予約だけでひとまず処理を終え、時間をおいて後に手動又は自動でリモートコピー開始要求を送信してもよい。

【0 0 4 4】

リモートコピー部 1 0 3 は、リモートコピー開始要求を受け取ると、そのリモートコピー開始要求からポートペアの IP アドレスを取り出し、通信プロトコル処理部 2 1 4 へそのポートペアを含むバッファ割り当て要求を通知する。通信プロトコル処理部 2 1 4 は、このバッファ割り当て要求を受信すると、後述するバッファ割り当て処理を行い、割り当てたバッファ 3 4 の先頭アドレスをリモートコピー部 1 0 3 へ通知する（ステップ 8 1 4）。リモートコピー部 1 0 3 は、このバッファアドレスを受け取ると、そのアドレスのバッファを用いて従来技術に従ってリモートコピー処理を行う（ステップ 8 1 5）。以上がリモートコピーを開始する時の処理のシーケンスである。

【0 0 4 5】

次に図 6 を用いて、バッファサイズ算出部 1 0 5 が行う推奨バッファサイズ算出処理の動作を説明する。バッファサイズ算出部 1 0 5 は、システム管理者がポートペア入力画面 6 0 1 を用いて入力したポートペアの IP アドレスをキーにして通信路情報テーブル 1 0 8 を検索する（ステップ 9 0 1）。次に検索にヒットしたレコードから、帯域 4 0 3、遅延 4 0 4、コネクション数 4 0 5、およびセッション数 4 0 6 を読み出す（ステップ 9 0 2）。次に帯域、遅延、コネクション数、セッション数をすべて乗じた結果を推奨バッファサイズとする（ステップ 9 0 3）。次にステップ 8 0 7 で得た使用可能バッファサイズをメモリ 2 2 3 から読み出し（ステップ 9 0 4）、推奨バッファサイズと比較する（ステップ 9 0 5）。使用可能バッファサイズが推奨バッファサイズより大きい場合、問題なく処理を終了する。使用可能バッファサイズが推奨バッファサイズよりも小さい場合、バッファ不足の警告を表示した後（ステップ 9 0 6）、処理を終了する。

【0 0 4 6】

次に図 7 を用いて、バッファ制御部 2 1 5 が行うバッファ予約処理の動作を説明する。バッファ制御部 2 1 5 は、まずバッファ予約要求を解析し、予約バッファサイズ、および使用可能バッファサイズを取り出し（ステップ 1 0 0 1）、両者を比較する（ステップ 1 0 0 2）。使用可能バッファサイズが予約バッファサイズ以上の場合、ステップ 1 0 0 6 へ移行する。逆に使用可能バッファサイズが予約バッファサイズより小さい場合、必要なサイズの未分割ページがあるか否かを調べる（ステップ 1 0 0 3）。ここで必要なサイズとは、予約バッファサイズから使用可能バッファサイズを引いた値である。また未分割ページとは、通信用メモリ領域 1 0 0 以外のメモリ上のページである。必要なサイズの未分割ページがある場合、ページをメモリバッファ単位に分割し、自由バッファリスト 3 5 に追加する（ステップ 1 0 0 5）。未分割ページのサイズが不足した場合、メモリ不足エラーを示すバッファ予約応答をストレージ管理装置 8 に返信する（ステップ 1 0 0 4）。

【0 0 4 7】

その後、バッファ制御部 2 1 5 は、自由バッファリスト 3 5 から必要量のメモリバッファを取り出し、予約バッファリスト 3 3 とする（ステップ 1 0 0 6）。次に当該ポートペアの IP アドレス、予約バッファリスト 3 3 の先頭メモリバッファのアドレス、および予約バッファサイズから成るレコードを予約バッファ情報テーブル 1 0 1 に追加し（ステップ 1 0 0 7）、処理を終了する。なお当該レコードがすでに存在している場合には、その予約バッファサイズを更新する。

【0 0 4 8】

次に図 8 を用いて、通信プロトコル処理部 2 1 4 が行うバッファ割り当て処理の動作を説明する。通信プロトコル処理部 2 1 4 は、まず受信したバッファ割り当て要求からポートペアの IP アドレスを取り出し、当該ポートペアをキーにして予約バッファ情報テーブル 1 0 1 を検索する（ステップ 1 1 0 1）。ヒットするレコードが存在した場合（ステップ 1 1 0 2 YES）、当該ポートペア間の通信用にバッファが予約されていると判断し、そのレコードからバッファアドレス 3 0 3 を取り出す。そしてそのバッファアドレスから始まる予約バッファリスト 3 3 から所定数のメモリバッファを取り出し、それらを連結したバッファ 3 4 を、要求元のアプリケーション、すなわちリモートコピー部 1 0 3 に割り当て（ステップ 1 1 0 3）、処理を終了する。

【0049】

なお割り当てたバッファ34について割当バッファ情報テーブル109に当該ポートペアのIPアドレス、バッファ34のバッファアドレスおよび割当バッファサイズを登録する。当該ポートペアについてすでにバッファ割当済であれば、予約バッファサイズ304のうち残りの予約バッファサイズからバッファ割当を行い、割当バッファ情報テーブル109のバッファアドレスおよび割当バッファサイズを更新する。ここで割当バッファサイズの上限は、予約バッファサイズであり、予約バッファサイズを越えるバッファ割当要求は拒否される。

【0050】

一方、検索にヒットするレコードが存在しない場合（ステップ1102 NO）、当該ポートペア間の通信にはバッファが予約されていないと判断し、自由バッファリスト35から所定数のメモリバッファを取り出し、それらを連結したバッファ34を、要求元のアプリケーションに割り当て（ステップ1104）、処理を終了する。なお割り当てたバッファ34について行う割当バッファ情報テーブル109のレコード登録又はレコード更新については、上記の通りである。なお予約バッファリスト33、および自由バッファリスト35から取り出すメモリバッファの数は、システム管理者があらかじめ通信インタフェース208ごとに設定する。

【0051】

本実施例では、ステップ814において、従来のリモートコピー処理を実行中には従来技術の方法によって、通信路の遅延や帯域の変化に応じて割り当てるバッファのサイズを変更することを想定している。

【0052】

また本実施例では、通信プロトコルとしてTCP/IPを想定しているが、本発明はFire Channelなど、ほかのプロトコルでも有効である。

【0053】

以上で本発明の第1の実施例を説明した。第1の実施例は、リモートコピーの開始以前にポートペア間でリモートコピー通信に必要なサイズのバッファを通信メモリ領域内に予約する。これによりホストとの通信の増加に起因する通信メモリ領域の不足を防ぎ、結果としてリモートコピーの中断やそれによる基幹業務の停止を防ぐことができる。

【0054】

次に第2の実施例について第1の実施例との相違点のみ説明する。第2の実施例は、第1の実施例と同様のリモートコピーシステムにおいて、リモートコピーの実行中に通信路情報の変化に応じてバッファサイズを更新する方法を説明する。

【0055】

図9は、ストレージ管理装置8が通信路情報管理装置3から通信路情報変更通知を受信した後、推奨バッファサイズを再計算し、マスタストレージ装置1が実行しているリモートコピーの予約バッファサイズを更新するときの処理手順を示す。

【0056】

マスタストレージ装置1があるポートペアの間にリモートコピーを実行中に、WAN 221の通信路情報を通信路情報取得装置4、5から収集している通信路情報管理装置3が、当該ポートペア間の帯域および（または）遅延の変化を検知すると、ストレージ管理装置8へ当該ポートペアのIPアドレス、ポートペア間の変化後の帯域、および遅延を含む通信路情報変更通知を送信する（ステップ1201）。このような帯域、遅延の変化の発生原因として、通信路における輻輳の発生、障害による経路の切り替えなどが挙げられる。

【0057】

ストレージ管理装置8が通信路情報変更通知を受信すると、情報管理部104は、その通信路情報変更通知からポートペアのIPアドレス、帯域、および遅延を取り出し、通信路情報テーブル108内の当該ポートペアを含むレコードを更新する。そして情報管理部104は、当該ポートペアを指定する通信インタフェース情報取得要求をマスタストレ

ジ装置 1 へ送信する（ステップ 1202）。マスタストレージ装置 1 が通信インタフェース情報取得要求を受信すると、通信インタフェース情報取得部 210 は、通信用メモリ領域 100 の使用可能バッファサイズを調べる。また通信インタフェース情報取得部 210 は、該当する割当バッファ情報テーブル 109 の該当するポートペアのレコードを参照し、その割当バッファサイズを取り出す（ステップ 1203）。そして通信インタフェース情報取得部 210 は、使用可能バッファサイズおよび割当バッファサイズを含む通信インタフェース情報取得応答を、ストレージ管理装置 8 へ送信する（ステップ 1204）。ストレージ管理装置 8 が通信インタフェース情報取得応答を受信すると、情報管理部 104 は、応答に含まれる使用可能バッファサイズおよび割当バッファサイズをメモリ 223 上に保持する。

【0058】

その後、バッファサイズ算出部 105 は、上記の推奨バッファサイズ算出処理を実行し、推奨バッファサイズを算出する（ステップ 1205）。なおここでは、ステップ 905 の比較式を、“推奨バッファサイズ－割当バッファサイズ＞使用可能バッファサイズ？”に置き換えて実行する。

【0059】

そして情報管理部 104 は、当該ポートペアおよび算出された推奨バッファサイズを含む予約バッファ更新要求をマスタストレージ装置 1 へ送信する（ステップ 1206）。マスタストレージ装置 1 がバッファ更新要求を受信すると、バッファ制御部 215 は、後述する予約バッファ更新処理を実行する（ステップ 1207）。バッファ制御部 215 は、予約バッファ更新処理を終了後、予約バッファ更新応答をストレージ管理装置 8 へ送信する（ステップ 1208）。

【0060】

次に図 10 を用いてバッファ制御部 215 の予約バッファ更新処理の動作を説明する。なおこの処理では、ステップ 1206 の予約バッファ更新要求で指定されるポートペアおよび推奨バッファサイズを入力値とする。バッファ制御部 215 は、まずポートペアの IP アドレスをキーにして予約バッファ情報テーブル 101 を検索する（ステップ 1301）。ヒットするレコードがない場合（ステップ 1302 NO）、図 7 のバッファ予約処理を実行し（ステップ 1303）、処理を終了する。

【0061】

ヒットするレコードがある場合（ステップ 1302 YES）、そのレコードから予約バッファのバッファアドレス 303 および予約バッファサイズ 304 を読み出す。そして推奨バッファサイズから予約バッファサイズを引いた値を更新バッファサイズとする（ステップ 1305）。

【0062】

更新バッファサイズが 0 より大きいとき（ステップ 1306 YES）、更新バッファサイズと使用可能バッファサイズを比較する。更新バッファサイズが使用可能バッファサイズ以上の場合（ステップ 1308 NO）、メモリ不足エラーを示すバッファ更新応答をストレージ管理装置 8 に返信し（ステップ 1309）、処理を終了する。逆に更新バッファサイズが使用可能バッファサイズより小さい場合（ステップ 1308 YES）、自由バッファリスト 35 から更新バッファサイズ分のメモリバッファを獲得し、それを予約バッファリスト 33 に連結し、予約バッファリスト 33 を拡大する（ステップ 1310）。

【0063】

一方、更新バッファサイズが 0 より小さい場合（ステップ 1306）、予約バッファリスト 33 から、更新バッファサイズ分のメモリバッファを取り出して自由バッファリスト 35 へ連結し、予約バッファリスト 33 を縮小する（ステップ 1307）。

【0064】

このような予約バッファリスト 33 のサイズ変更後、予約バッファ情報テーブル 101 の予約バッファサイズ 304 の値を推奨バッファサイズに更新し（ステップ 1311）、処理を終了する。

【0065】

以上で第2の実施例について説明した。第2の実施例では、ストレージ管理装置が、ポートペア間の通信路の遅延や帯域の変化に応じて、マスタストレージ装置1の予約バッファのサイズを変更することにより、リモートコピー実行中に、輻輳や障害発生に起因する経路切り替えが発生したときのバッファ不足を防ぐことができる。

【0066】

次に第3の実施例について第1の実施例および第2の実施例との相違点のみ説明する。第3の実施例は、第1の実施例と同様のリモートコピーシステムにおいて、ポートペアとリモートサイトのポート番号の組に対してバッファを予約することを可能にする。

【0067】

ここでポート番号とは、IPアドレスに対して付加されるサブアドレスであり、TCP/IP通信を行う際に、そのIPアドレスのリモートストレージ装置のどのアプリケーション層のプロトコルを使用した通信を行うかを示す番号である。例えばポートペアとリモートサイトのポート番号の組とは、マスタポートIPアドレス「211.19.1.20」、リモートポートIPアドレス「133.185.12.20」、およびリモートサイトのポート番号「25」から成る。

【0068】

まず第3の実施例で用いるデータ構成について、相違点のみ説明する。本実施例では、図3(a)の予約バッファ情報テーブル101、割当バッファ情報テーブル109および図3(b)の通信路情報テーブル108に、リモートサイトのポート番号を格納するカラムを追加する。

【0069】

次に第3の実施例に係わるGUIについて、相違点のみ説明する。本実施例では、GUI制御部107が、図11に示す予約対象入力画面1401を表示する。予約対象入力画面1401は、マスタポートのIPアドレスを入力する領域1402、リモートポートのIPアドレスを入力する領域1403、リモートサイトのポート番号を入力する領域1404、ポートペアとリモートサイトのポート番号の組当たりのセッション数を入力する領域1405および1セッション当たりのコネクション数を入力する領域1406を設ける。また予約対象入力画面1401は、これらのパラメータの登録を指示するためのOKボタン1407および入力を取り消すよう指示するためのキャンセルボタン1408を有する。また図4(b)は、第1の実施例と同様である。

【0070】

次に第3の実施例に係わる処理手順、および第1の実施例あるいは第2の実施例との処理上の相違点について説明する。

【0071】

まず図5の処理手順について、第1の実施例との相違点について説明する。第3の実施例は、図5のステップ801において、予約対象入力画面1401を表示する。ステップ802からステップ809までは、第1の実施例と同様である。

【0072】

ステップ810において、バッファサイズ送信部106は、マスタストレージ装置1へ送信するバッファ予約要求で指定する情報としてリモートサイトのポート番号を加える。ステップ811およびステップ812は、第1の実施例と同様である。

【0073】

ステップ813において、情報管理部104は、リモートコピー部103へ通知するリモートコピー開始要求で指定する情報としてリモートサイトのポート番号を加える。ステップ814において、リモートコピー部103は、受信したリモートコピー開始要求から取り出す情報、および通信プロトコル処理部214へ通知するバッファ割当要求で指定する情報として、リモートサイトのポート番号を加える。それ以外は、第1の実施例と同様である。ステップ815は、第1の実施例と同様である。以上で図5の処理手順について相違点のみ説明した。

【0074】

次に図6の推奨バッファサイズ算出処理について、第1の実施例との相違点について説明する。第3の実施例では、図6のステップ901において、バッファサイズ算出部105が、通信路情報テーブル108を検索するキーとして用いる情報としてリモートサイトのポート番号を加える。ステップ902以降は、第1の実施例と同様である。

【0075】

次に図7のバッファ予約処理について、第1の実施例との相違点について説明する。第3の実施例のステップ1001から1006は、第1の実施例と同様である。ステップ1007において、バッファ制御部215は、予約バッファ情報テーブル101に追加するレコードの情報にリモートサイトのポート番号を含める。

【0076】

次に図8のバッファ割当処理について、第1の実施例との相違点について説明する。第3の実施例では、図8のステップ1101において、通信プロトコル制御部214は、受信したバッファ割当要求から取り出す情報、および予約バッファ情報テーブル101を検索するキーとして用いる情報に、リモートサイトのポート番号を加える。ステップ1102から1104までは、割当バッファ情報テーブル109に登録する情報にリモートサイトのポート番号を含めること、および「ポートペア」を「ポートペアとリモートサイトのポート番号の組」と読み替えること以外は、第1の実施例と同様である。

【0077】

次に図9の処理手順について第2の実施例との相違点について説明する。第3の実施例のステップ1201は、第2の実施例と同様である。

【0078】

ステップ1202から1208までにおいて、まず情報管理部104は、受信した通信路情報変更通知からポートペア、帯域および遅延を取り出し、そのポートペアをキーにして通信路情報テーブル108を検索し、ヒットした全てのレコードの帯域403と遅延404とを、通信路情報変更通知から取り出した帯域と遅延とで上書きする。次にヒットした各レコードについて、以下の処理を行う。まず情報管理部104は、通信インタフェース情報取得要求をマスタストレージ装置1に送信する（ステップ1202）。この通信インタフェース情報取得要求は、ポートペアとリモートサイトのポート番号を指定する。マスタストレージ装置1が通信インタフェース情報取得要求を受信すると、通信インタフェース情報取得部210は、通信用メモリ領域100の使用可能バッファサイズを調べる。また通信インタフェース情報取得部210は、該当する割当バッファ情報テーブル109の該当するレコードを参照し、その割当バッファサイズを取り出す（ステップ1203）。そして通信インタフェース情報取得部210は、通信インタフェース情報取得応答をストレージ管理装置8へ送信する（ステップ1204）。ストレージ管理装置8が通信インタフェース情報取得応答を受信すると、バッファサイズ算出部105は、上記の推奨バッファサイズ算出処理を実行し、推奨バッファサイズを得る（ステップ1205）。そして情報管理部104は、ポートペア、リモートサイトのポート番号、および推奨バッファサイズを含む予約バッファ更新要求を、マスタストレージ装置1へ送信する（ステップ1206）。マスタストレージ装置1がバッファ更新要求を受信すると、バッファ制御部215は、後述する予約バッファ更新処理を実行する（ステップ1207）。バッファ制御部215は、予約バッファ更新処理を終了後、予約バッファ更新応答をストレージ管理装置8へ送信する（ステップ1208）。以上のステップ1202から1208までの処理を、上記の検索にヒットした各レコードについて繰り返す。

【0079】

次に図10の予約バッファ更新処理について、第2の実施例との相違点について説明する。第3の実施例のステップ1301において、バッファ制御部215は、予約バッファ情報テーブル101を検索するキーとして用いる情報に、リモートサイトのポート番号を加える。ステップ1301から1311は、「ポートペア」を「ポートペアとリモートサイトのポート番号の組」と読み替える以外、第1の実施例と同様である。

【0080】

以上で第3の実施例について説明した。第3の実施例は、バッファの予約対象をポートペアとリモートサイトのポート番号の組とした。これによって予約したバッファを利用可能なコネクションを、リモートサイトのポート番号で識別可能なアプリケーション層のプロトコルを使うコネクションに限定することができる。ポートペア間で様々なプロトコルを使う通信を行う環境において特に有効である。

【0081】

次に第4の実施例について第1および第2の実施例との相違点のみ説明する。第4の実施例は、第1の実施例と同様のリモートコピーシステムにおいて、コネクションについてバッファを予約することを可能にする。

【0082】

まず第4の実施例で用いるデータの形式について、相違点のみ説明する。図12は、マスタストレージ装置1のリモートコピー部103が、予約したバッファを使用するときに識別するバッファIDを登録するための予約バッファID情報テーブル1500のデータ構成を示す。予約バッファID情報テーブル1500は、マスタストレージ装置1のメモリ218に格納される。予約バッファID情報テーブル1500は、配列構造を成し1以上のレコードを格納可能である。予約バッファID情報テーブル1500の各レコードは、マスタポートIPアドレス1501、リモートポートIPアドレス1502、およびバッファID1503から構成される。例えばレコード1504は、マスタポートIPアドレス「211. 19. 1. 20」とリモートポートIPアドレス「133. 185. 12. 20」間のコネクションについて、予約したバッファのバッファIDが「10000」であることを示す。

【0083】

また図3(a)の予約バッファ情報テーブル101にバッファIDを格納するカラムを追加する。割当バッファ情報テーブル109、および図3(b)の通信路情報テーブル108は、第1の実施例と同様である。

【0084】

次に第4の実施例に係わるGUIについて相違点のみ説明する。本実施例のGUI制御部107は、図4(a)のポートペア入力画面601を表示する際に領域604と領域605、およびそれらの左側の文字列を表示しない。また図4(b)のサイズ指定画面701は、第1の実施例と同様である。

【0085】

次に第4の実施例に係わる処理手順および動作について、相違点のみ説明する。まず図5の処理手順について、第1の実施例との相違点について説明する。

【0086】

第4の実施例のステップ801において、GUI制御部107は、上記のとおり、図4(a)のポートペア入力画面601を表示する。さらに情報管理部104は、通信路情報テーブル108にレコードを追加する際、セッション数を1、コネクション数を1とする。ステップ802から810は、第1の実施例と同様である。

【0087】

ステップ811において、バッファ制御部215は、バッファを予約できた場合、バッファIDを発行する。そしてステップ812において、バッファ制御部215は、バッファIDを指定するバッファ予約応答を、ストレージ管理装置8へ送信する。このバッファ予約応答を受信したリモートコピー部103は、予約バッファID情報テーブル1500に、当該ポートペアのIPアドレスとバッファIDから成るレコードを追加する。ステップ813は、第2の実施例と同様である。

【0088】

ステップ814において、リモートコピー部103は、まず受信したリモートコピー開始要求からポートペアを取り出し、それをキーとして予約バッファID情報テーブル1500を検索する。そしてリモートコピー部103は、ヒットしたレコードからバッファID

D 1 5 0 3 を取り出し、そのバッファ I D を指定するバッファ割当要求を通信プロトコル処理部 2 1 4 へ通知する。その後は、第 1 の実施例と同様である。ステップ 8 1 5 は、第 1 の実施例と同様である。

【 0 0 8 9 】

図 6 の推奨バッファサイズ算出処理は、第 1 の実施例と同様である。

【 0 0 9 0 】

次に図 7 のバッファ予約処理について、第 1 の実施例との相違点について説明する。第 4 の実施例のステップ 1 0 0 1 から 1 0 0 6 は、第 1 の実施例と同様である。ステップ 1 0 0 7 において、バッファ制御部 2 1 5 は、まずバッファ I D を発行し、次に予約バッファ情報テーブル 1 0 1 に追加するレコードの情報に、バッファ I D を含める。

【 0 0 9 1 】

次に図 8 のバッファ割当処理について、第 1 の実施例との相違点について説明する。第 4 の実施例のステップ 1 1 0 1 において、バッファ制御部 2 1 5 は、予約バッファ情報テーブル 1 0 1 を検索するキーとして用いる情報に、バッファ I D を用いる。ステップ 1 1 0 2 から 1 1 0 4 までは、第 1 の実施例と同様である。

【 0 0 9 2 】

次に図 9 の処理手順については、第 2 の実施例と同様である。

【 0 0 9 3 】

次に図 1 0 のバッファ更新処理について、第 2 の実施例との相違点について説明する。ステップ 1 3 0 1 において、バッファ制御部 2 1 5 は、バッファ更新要求から取り出したポートペアをキーにして予約バッファ情報テーブル 1 0 1 を検索した際に、ヒットするレコードが複数あり得る。その場合、ヒットしたレコードそれぞれに対して、ステップ 1 3 0 4 から 1 3 1 1 までを実行する。ステップ 1 3 0 4 から 1 3 1 1 までは、第 2 の実施例と同様である。

【 0 0 9 4 】

以上で第 4 の実施例について説明した。第 4 の実施例は、コネクションをバッファの予約対象とすることにより、予約したバッファを特定のコネクションが専有できるようになる。

【 0 0 9 5 】

なお第 4 の実施例において、アプリケーションは、バッファ I D で識別される予約バッファを、任意の 2 以上のコネクションで共用してもよい。例えばマスタストレージ装置 1 の複数のポートとリモートストレージ装置 2 の複数のポートの間で、複数のコネクションを用いてリモートコピーを実行する場合に、バッファ I D で識別される予約バッファをこれらのコネクションで共用するなどの応用がある。

【 0 0 9 6 】

なお第 1、第 2、第 3 および第 4 の実施例において、リモートコピー部 1 0 3 は、図 5 のリモートコピー処理 8 1 5 でリモートコピー処理を行うときに、この処理で使用するコネクションについて輻輳ウィンドウの最大値を予約したバッファのサイズと同じ値とするように、通信プロトコル処理部 2 1 4 に対して指示してもよい。輻輳ウィンドウについては、R F C 2 5 8 1 に記載されている。

【 0 0 9 7 】

また図 1 0 のステップ 1 3 1 1 における予約バッファサイズの更新後、バッファ制御部 2 1 5 は、輻輳ウィンドウの最大値を予約バッファサイズの更新値に変更するように、通信プロトコル処理部 2 1 4 に指示してもよい。これにより、通信路の性質に応じた適切な輻輳ウィンドウの最大値を設定して、通信路の性能に比べて輻輳ウィンドウが小さ過ぎる場合の通信のスループットの低下を防ぐことができる。

【 0 0 9 8 】

なお第 1、第 3 および第 4 の実施例において、3 種類の予約対象を示しているが、システム管理者が予約バッファを用いるアプリケーションの形態に応じて予約対象を選択可能にしてもよい。例えばストレージ管理装置 8 の G U I 制御部 1 0 7 が、ポートペアなどの

予約対象入力画面を表示する前に、図13に示す予約対象選択画面1601を表示することによって実現できる。

【0099】

図13は、リモートコピーシステムの管理者が、予約バッファを割り当てる対象を選択するための予約対象選択画面1601の表示例である。ストレージ管理装置8のGUI制御部107は、予約対象選択画面1601をディスプレイ227に表示する。システム管理者は、この画面を介してポインティング装置229を用いて予約対象を設定する。

【0100】

予約対象選択画面1601は、予約対象として、ポートペアを選択するポートペア選択ボタン1602、ポートペアおよびリモートサイトのポート番号を選択するボタン1603およびコネクションを選択するコネクション選択ボタン1604を設ける。また予約対象選択画面1601は、これらのボタンのうちシステム管理者が選択した予約対象を決定するよう指示するためのOKボタン1605および予約対象の選択を取り消すよう指示するためのキャンセルボタン1606を有する。

【0101】

以下、処理動作の一例について説明する。まずステップ801において、ストレージ管理装置8のGUI制御部107は、システム管理者からリモートコピー開始操作を受け付けると、ディスプレイ227に図13に示す予約対象選択画面1601を表示する。システム管理者は、ポインティング装置229を用いて、ボタン1602、1603、1604のいずれか一つを選択する。システム管理者は、選択後、ボタン1605もしくはボタン1606を押下する。システム管理者がポートペア選択ボタン1602を選択した場合、以後第1の実施例で説明した処理を実行する。システム管理者がポートペアおよびリモートサイトのポート番号選択ボタン1603を選択した場合、以後第3の実施例で説明した処理を実行する。システム管理者がコネクション選択ボタン1604を選択した場合、以後第4の実施例で説明した処理を実行する。

【0102】

なお上記の3つ以外の選択対象を追加してもよい。例えば様々な通信を行うアプリケーション、リモートコピーの対象となるマスタサイトとリモートサイトの論理ボリュームのペア、またマスタサイトの複数の論理ボリュームのグループとリモートサイトの複数の論理ボリュームのグループのペアなどがある。

【0103】

次に第5の実施例について説明する。第5の実施例は、第1の実施例と同様のリモートコピーシステムにおいて、リモートコピーの実行中にストレージ装置のキャッシュメモリの不足に応じてバッファサイズを更新する方法を説明する。

【0104】

第5の実施例において、マスタストレージ装置1のリモートコピー部103は、リモートコピーの実行中に、キャッシュメモリ102の空き領域サイズを定期的に収集する。キャッシュメモリの空き領域サイズが警告しきい値以下になると、リモートコピー部103は、リモートコピーに使用中のポートペアをキーにして予約バッファ情報テーブル101を検索し、ヒットしたレコードから予約バッファサイズ304を読み出す。ここで警告しきい値とは、システム管理者がストレージ管理装置8を用いて指定したものである。あるいはリモートコピー部103は、起動時に各種設定情報を記述した構成定義ファイルを読み出し、その中から警告しきい値を取り出してもよい。

【0105】

そしてリモートコピー部103は、ポートペアと予約バッファサイズとを指定するバッファ制御自動化要求を、ストレージ管理装置8の情報管理部104へ送信する。ストレージ管理装置8の情報管理部104は、受信したバッファ制御自動化要求からポートペアと予約バッファサイズを取り出し、図5のステップ802からステップ808までを実行し、当該ポートペアについての推奨バッファサイズを得る。そして情報管理部104は、この推奨バッファサイズと予約バッファサイズとを指定する予約バッファ更新要求をマスタ

ストレージ装置 1 のバッファ制御部 215 へ送信する。バッファ制御部 215 は、予約バッファ更新要求を受信すると、図 10 の予約バッファ更新処理を実行する。

【0106】

以上で第 5 の実施例について説明した。第 5 の実施例は、リモートコピーの実行中にキャッシュメモリの空き領域がしきい値を下回るキャッシュメモリ不足が起こった場合、リモートコピーが使用している予約バッファのサイズを更新する。キャッシュメモリ不足が管理者による手動で過小なバッファを予約したことに起因する場合、本実施例によって予約バッファが拡大してデータの送信量を増大し、キャッシュメモリ不足を回避でき、結果としてキャッシュメモリの飽和に伴う基幹業務の停止を防ぐことができる。

【0107】

上記実施例によれば、リモートコピーの開始以前に、ポートペア間のリモートコピー通信に必要なサイズのバッファを、通信用メモリ領域内に予約することにより、ホストとの通信の増加に起因する通信用メモリ領域の不足を防ぎ、結果としてリモートコピーの中断やそれによる基幹業務の停止を防ぐことができる。

【0108】

また上記実施例によれば、ストレージ管理装置が、ポートペア間の通信路の遅延や帯域の変化に応じて、マスタストレージ装置のバッファのサイズを変更することにより、リモートコピー実行中に、輻輳や障害発生に起因する経路切り替えが発生したときのバッファ不足を防ぐことができる。

【0109】

また上記実施例によれば、システム管理者が、予約バッファを割り当てる対象を、ポートペア、ポートペアおよびリモートサイトのポート番号およびコネクションのいずれかを選択可能とすることにより、リモートコピーの形態に応じた適切なバッファ予約を行うことができる。

【0110】

また上記実施例によれば、キャッシュメモリが不足した際に、予約済みバッファおよび輻輳ウィンドウを拡大してデータの送信量を増大することにより、キャッシュメモリからのデータの消去速度を上げ、結果としてキャッシュメモリの飽和に伴う基幹業務の停止を防ぐことができる。

【図面の簡単な説明】

【0111】

【図 1】 第 1 の実施例のリモートコピーシステムの構成図である。

【図 2】 実施例の通信用メモリ領域のデータ構成図である。

【図 3】 実施例の予約バッファ情報テーブルおよび通信路情報テーブルのデータ構成図である。

【図 4】 ポートペア入力画面およびサイズ指定画面の表示例を示す図である。

【図 5】 第 1 の実施例のリモートコピー実行までの処理手順を示す図である。

【図 6】 実施例の推奨バッファサイズ算出処理の手順を示すフローチャートである。

【図 7】 実施例のバッファ予約処理の手順を示すフローチャートである。

【図 8】 実施例のバッファ割り当て処理の手順を示すフローチャートである。

【図 9】 第 2 の実施例のバッファサイズ更新までの処理手順を示す図である。

【図 10】 第 2 の実施例のバッファ更新処理の手順を示すフローチャートである。

【図 11】 第 3 の実施例の予約対象入力画面の表示例を示す図である。

【図 12】 第 4 の実施例の予約バッファ ID 情報テーブルのデータ構成図である。

【図 13】 予約対象選択画面の表示例を示す図である。

【符号の説明】

【0112】

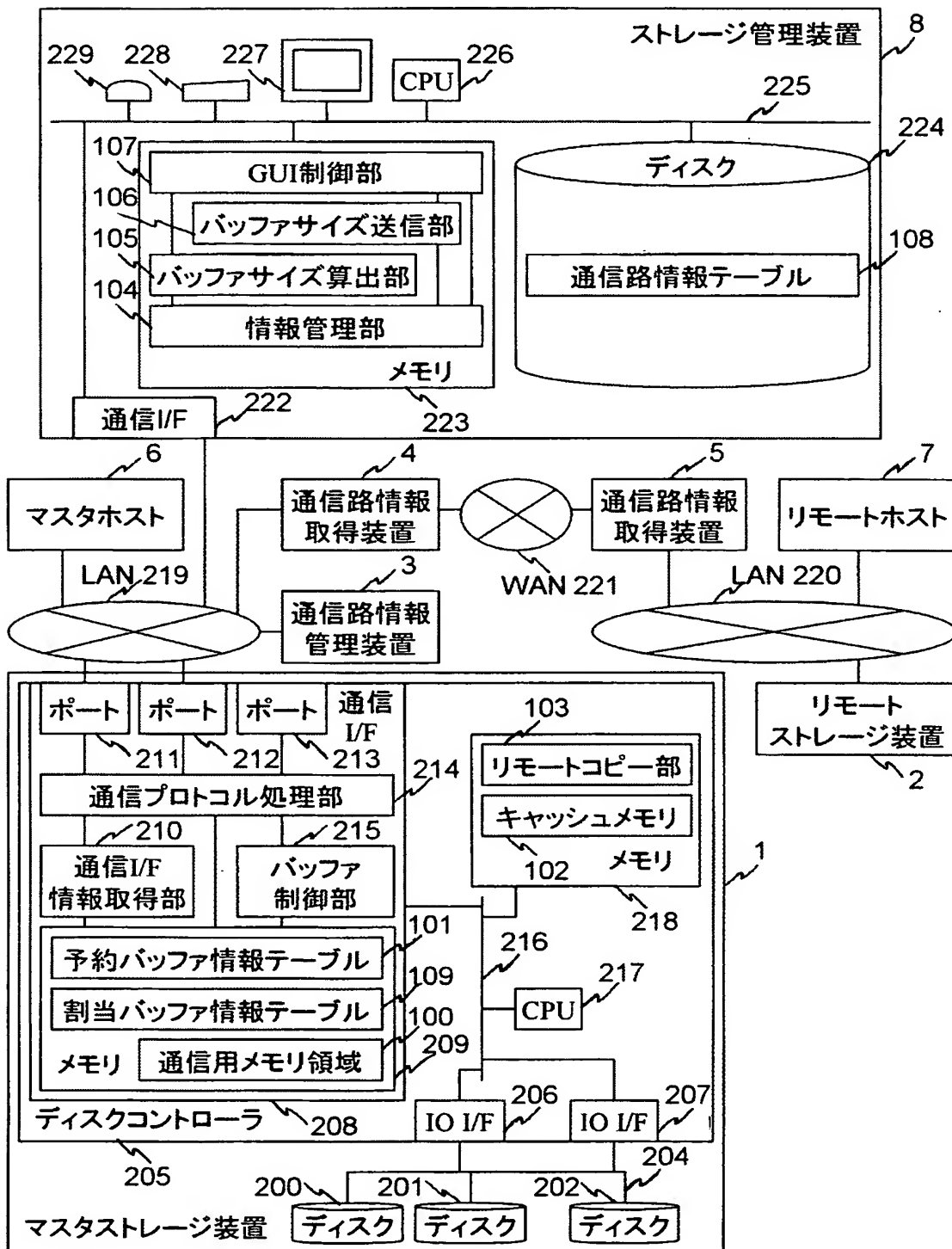
1…マスタストレージ装置、2…リモートストレージ装置、3…通信路情報管理装置、6…マスタホスト、7…リモートホスト、8…ストレージ管理装置、100…通信用メモリ領域、101…予約バッファ情報テーブル、103…リモートコピー部、104…情報

管理部、1 0 5…バッファサイズ算出部、1 0 6…バッファサイズ送信部、1 0 8…通信
路情報テーブル、1 0 9…割当バッファ情報テーブル、2 1 0…通信インタフェース情報
取得部、2 1 5…バッファ制御部。

【書類名】図面

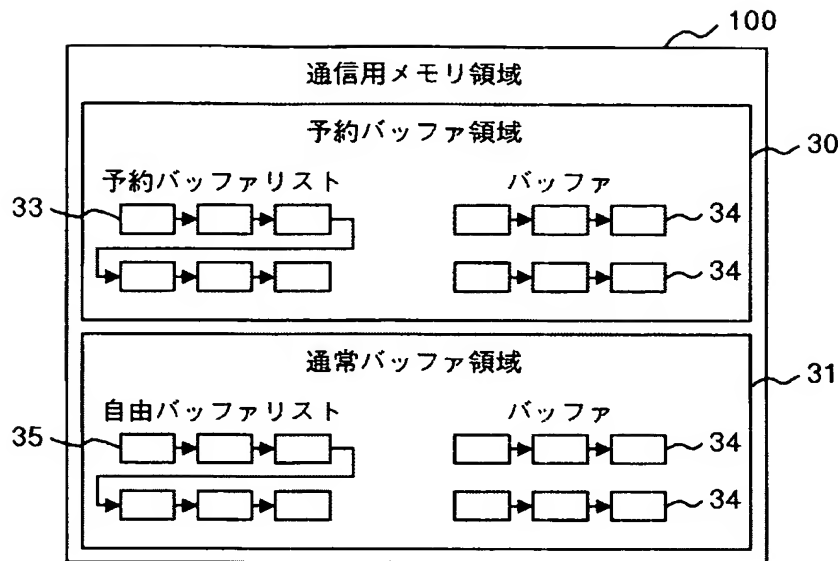
【図 1】

図 1



【図 2】

図 2



【図 3】

図 3

(a) 101 予約バッファ情報テーブル

	301	302	303	304
	マスタポート IP アドレス	リモートポート IP アドレス	バッファアドレス	予約バッファ サイズ
305	211.19.1.20	133.185.12.20	0x0FFFFFFF	2.1M
	211.19.3.85	172.83.1.21	0xFF00CC	4.5M

(b) 108 通信路情報テーブル

	401	402	403	404	405	406
	マスタポート IP アドレス	リモートポート IP アドレス	帯域	遅延	セッション数	コネクション数
407	211.19.1.20	133.185.12.20	10M	2.8 μ	2	3
	211.19.3.85	172.83.1.21	45M	1.2 μ	1	4

【図 4】

図 4

(a)

ポートペア入力画面

マスタポート IP アドレス 602

リモートポート IP アドレス 603

ポートペアあたりのセッション数 604

1セッション当りのコネクション数 605

OK 606 キャンセル 607

(b)

サイズ指定画面

使用可能なバッファサイズ 702

推奨バッファサイズ 703

704 ○ 自動指定

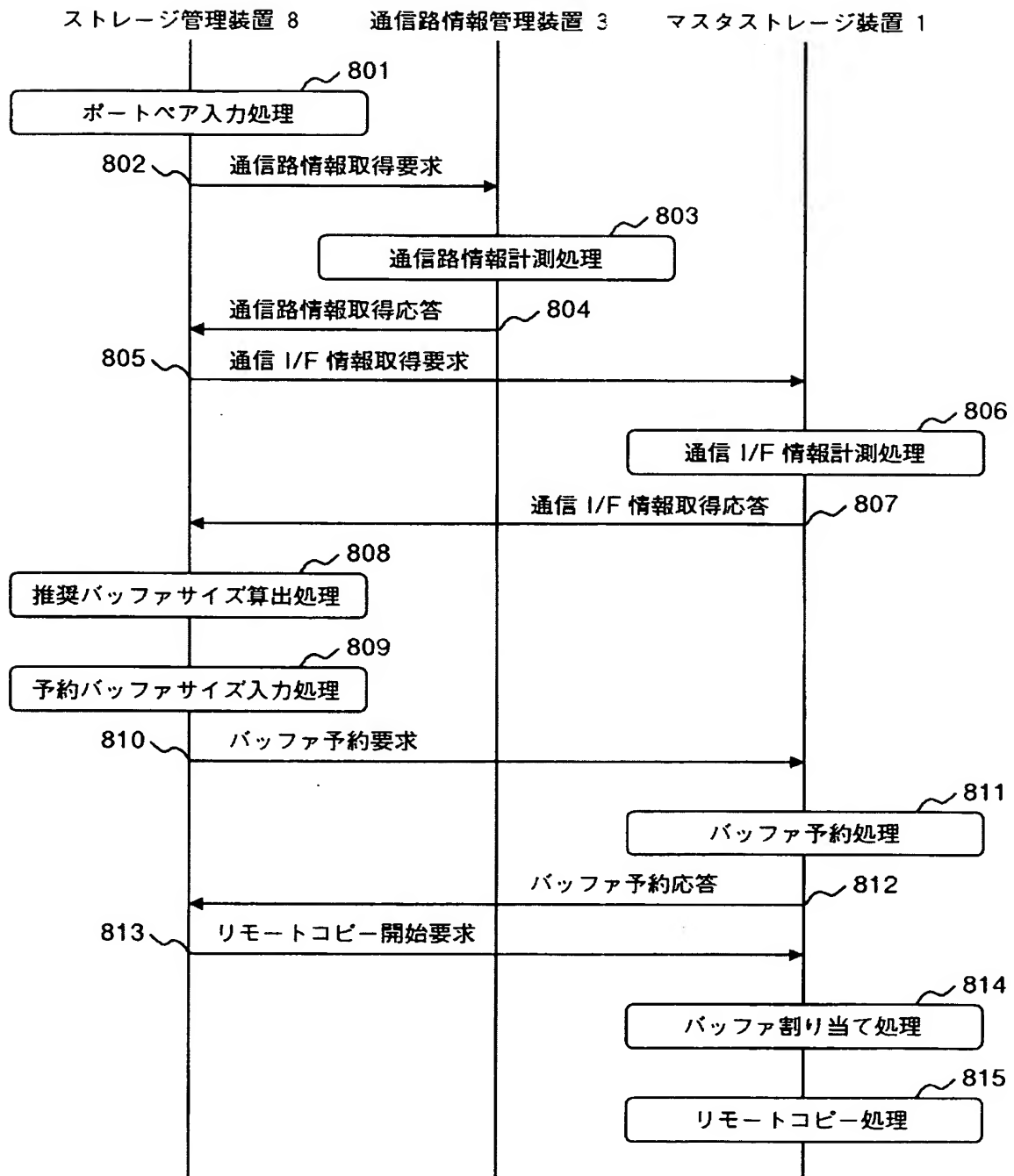
705 ● 手動指定

予約バッファサイズの入力 706

OK 707 キャンセル 708

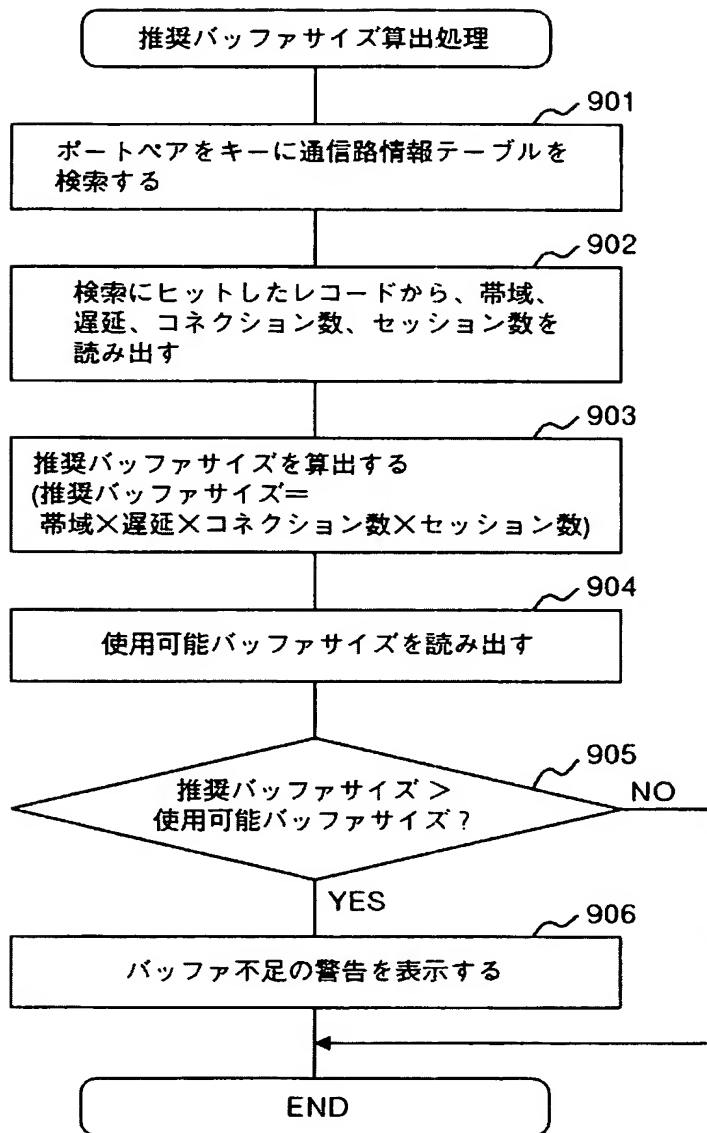
【図 5】

図 5



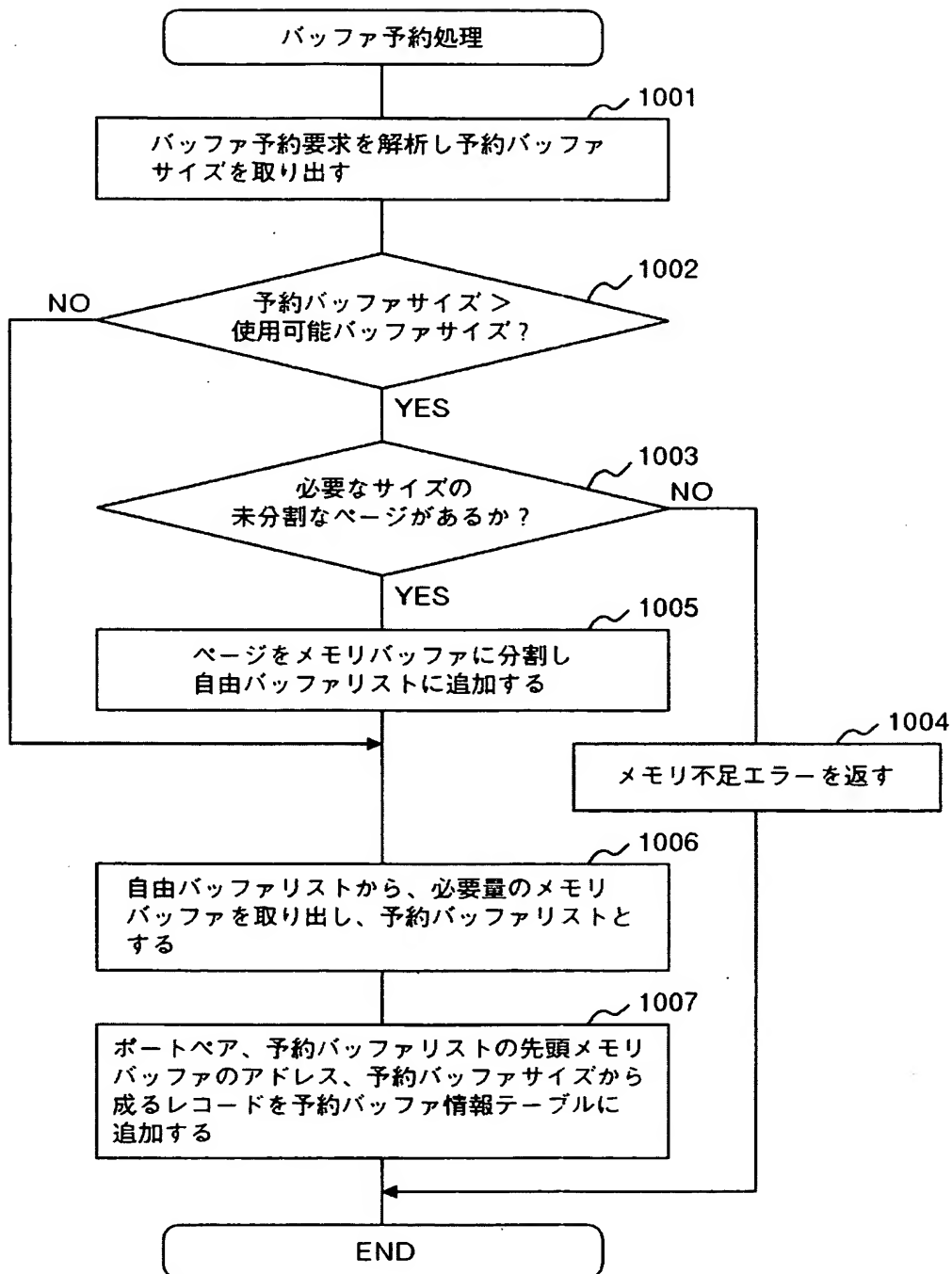
【図 6】

図 6



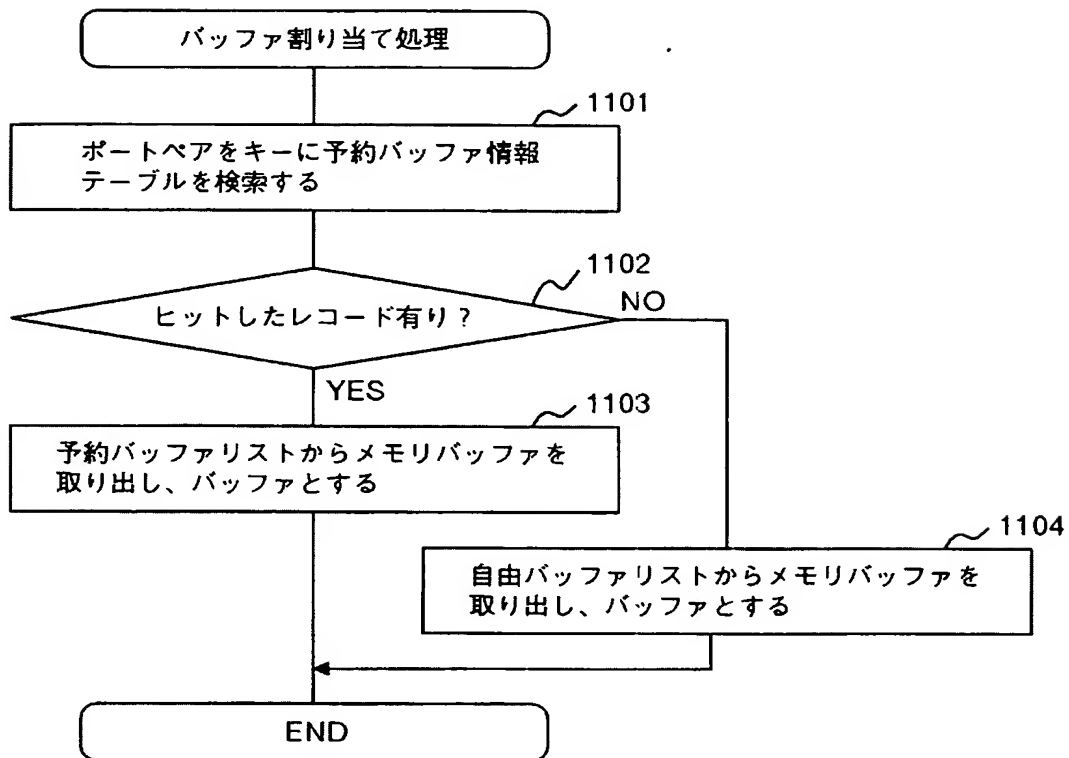
【図 7】

図 7



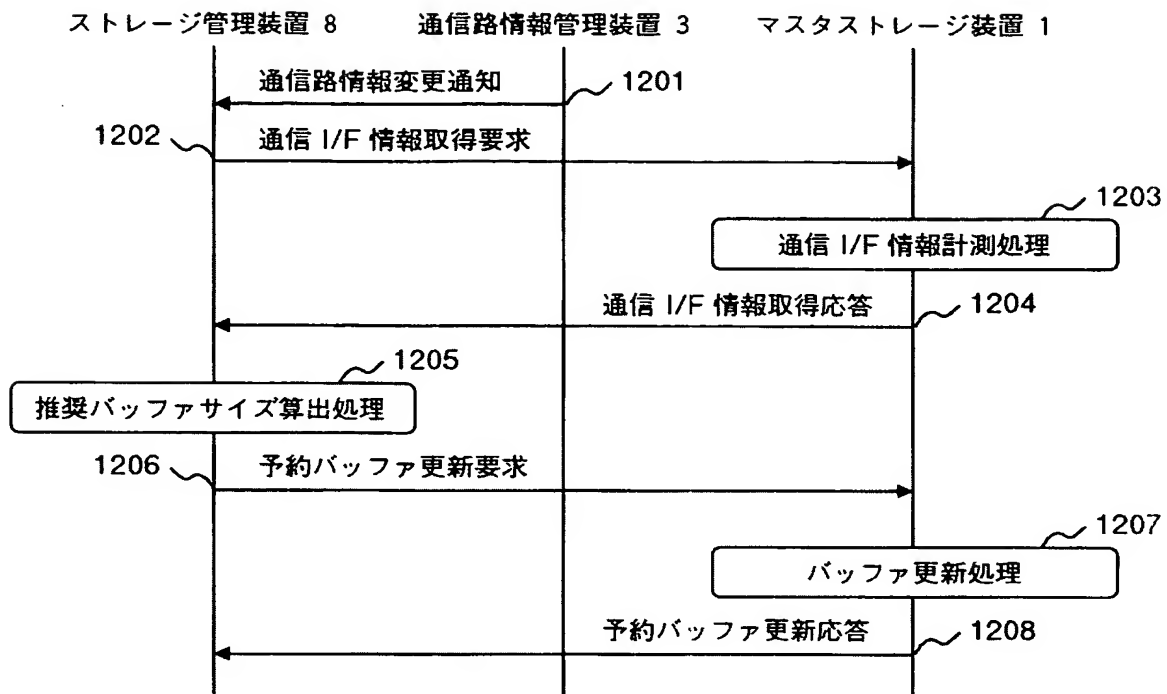
【図 8】

図 8



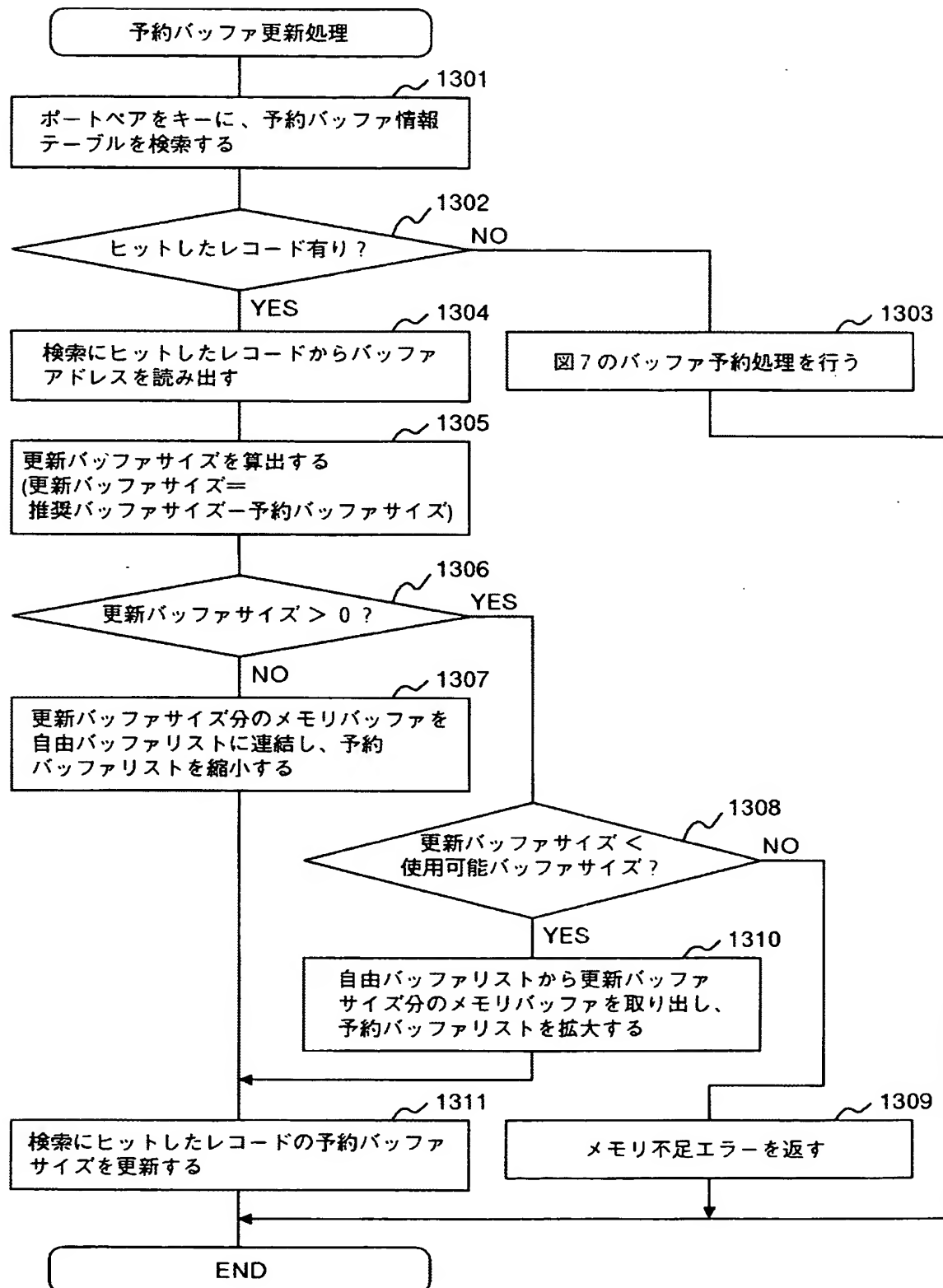
【図 9】

図 9



【図 10】

図 10



【図 11】

図 11

予約対象入力画面

マスタポート IP アドレス 1402

リモートポート IP アドレス 1403

リモートサイトのポート番号 1404

ポートペアとリモートサイトのポート
番号の組あたりのセッション数 1405

1セッションあたりのコネクション数 1406

OK 1407 キャンセル 1408

【図 12】

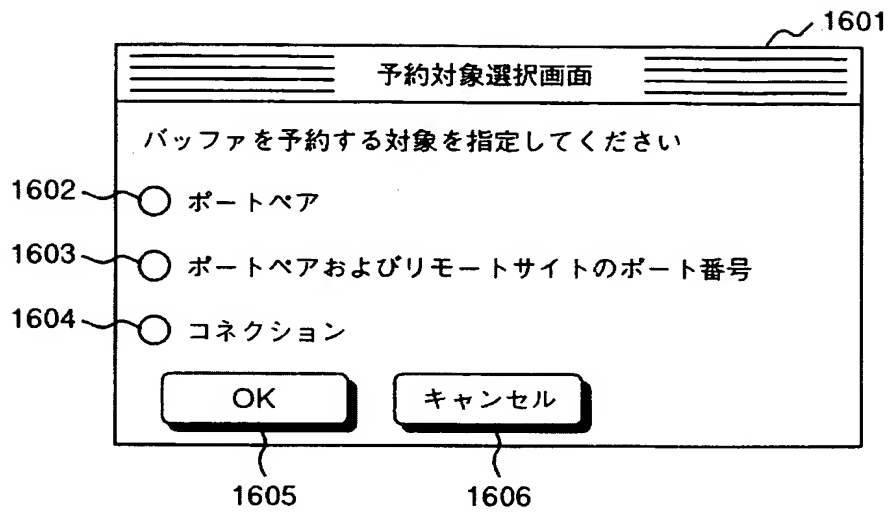
図 12

1500 予約バッファ ID 情報テーブル

1501			1502	1503
マスタポート IP アドレス		リモートポート IP アドレス	バッファ ID	
211.19.1.20		133.185.12.20	10000	
211.19.3.85		172.83.1.21	15000	
...		

【図 13】

図 13



【書類名】 要約書**【要約】**

【課題】 ネットワークを介してホストおよび他のストレージ装置と通信するストレージ装置において、ストレージ装置間のリモートコピーのためのバッファを確保する。

【解決手段】 マスタストレージ装置 1 の通信用メモリ領域 1 0 0 は、複数のメモリバッファから成る使用可能バッファと、通信のために割り当てられた複数のメモリバッファから成る使用中バッファとを有する。バッファ制御部 2 1 5 は、ストレージ管理装置 8 からのバッファ予約要求に応じて使用可能バッファから指定された予約バッファサイズ分の予約バッファを確保し、ストレージ管理装置 8 からのリモートコピー開始要求に応答して予約バッファ中のメモリバッファを割り当てて使用中バッファとする。リモートコピー部 1 0 3 は、使用中バッファを用いてリモートストレージ装置 2 との間でリモートコピーを実行する。

【選択図】 図 1

特願 2 0 0 3 - 3 5 3 2 1 2

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 5 1 0 8]

1. 変更年月日 1 9 9 0 年 8 月 3 1 日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台 4 丁目 6 番地

氏 名 株式会社日立製作所